Crystallographic Structure Determination of Bacteriophage Endolysins

Marta Sanz-Gaitero[1,2,3] and Mark J. van Raaij[1*]

[1]Centro Nacional de Biotecnologia, CSIC, Calle Darwin 3, E-28049 Madrid, Spain.
[2]Cork Institute of Technology, Bishopstown, Cork T12 P928, Ireland.
[3]Current address: Centre for Molecular Medicine Norway, University of Oslo, Gaustadalléen 21, 0349 Oslo, Norway.

*Corresponding author: mjvanraaij@cnb.csic.es

DOI: https://doi.org/10.21775/cimb.040.165

Abstract

Bacteriophages produce endolysins that target and cleave the hosts peptidoglycan to release their progeny at the end of the infection cycle. These proteins can be used for the eradication of pathogenic bacteria, but also for their detection. Endolysins may contain a single catalytic domain or several domains, including a cell wall binding domain. To understand their function in detail and design mutated or chimeric molecules with novel properties, knowledge of their structures and detailed mechanisms is necessary. X-ray protein crystallography is an excellent method to obtain high-resolution structures of biological macromolecules, and here we describe the method and the folds of known endolysin domains.

Introduction

Bacteriophages (phages) are a promising treatment for antibiotic-resistant pathogens (Coffey and Buttimer; 2020). However, they also present some

drawbacks. A safety concern is their possible impact on non-target microbiota. Phages can modify their bacterial targets causing them to express virulence genes or transduce DNA between bacteria, although these risks can be reduced by avoiding the use of temperate phages. As they are unknown entities to the host, they may induce both adaptive and innate immunological reactions (Majewska et al., 2019). Phage formulations may also contain harmful products, such as endotoxins, due to incomplete purification. Specific self-replication of phages at the infection site is a major advantage. However, a self-replicating entity may change its properties by spontaneous mutation or recombination with host sequences such as prophages. Self-replication also means that the pharmacological dose is difficult to calculate (Coffey and Buttimer; 2020). To avoid these drawbacks, in some cases, it may be better to use isolated phage proteins.

At the end of the infection cycle, once the phage progeny has been generated, the new virions need to be released from the bacteria. To this end, phages generate proteins called endolysins, able to break peptidoglycan links. This causes the bacteria to burst, due to the osmotic pressure. However, to reach the peptidoglycan layer, endolysin molecules first have to cross the cytoplasmic membrane. To allow this passage, phages produce pore-generating proteins called holins (Young, 2013), which insert themselves in the membrane and generate pores, through which the endolysins pass. Endolysins may have a single domain, with a catalytic activity. Other endolysins are multi-modular, combining cell wall binding domains and catalytic domains. Different catalytic domains cleave different chemical bonds of the peptidoglycan: muramidases, glycosidases, peptidases or amidases (Loessner, 2005).

Endolysins can cause the lysis of Gram-positive bacteria when applied exogenously (Fischetti, 2008; Coffey and Buttimer; 2020). In the case of Gram-negative bacteria, endolysins can be genetically engineered to allow them to pass through the outer membrane and lyse their host, for example by adding an outer-membrane-permeabilizing peptide (O'Sullivan et al., 2016; Coffey and Buttimer; 2020).

Endolysin applications are varied. They can be used as antimicrobial agents, for bacterial identification or for surface sterilization. Each of these applications requires the use of endolysins with different characteristics. For some applications, a wide lytic range me be necessary, for others a high specificity may be desired. Resistance to different temperature, pH, salinity or humidity ranges may also be required.

To obtain custom-designed endolysins, several techniques can be applied. Mutagenesis and protein truncation can be used to modify the lytic and antibacterial activity, cell-wall binding domain dependence, specificity, half-life and thermostability of endolysins. Recombining domains or domain swapping has been used successfully to create endolysins with altered properties (chimeolysins), such as activity against a certain type of bacteria or increased solubility (Yang et al., 2014; Gerstmans et al., 2018). Sequences can be added to these domains to increase their solubility or hydrophobicity, as in the case of artilysins, where an outer membrane permeabilizing region is added to allow them to pass through the bacterial outer membrane, allowing them to be effective even when applied exogenously. Cell wall binding domains of endolysins have been fused to fluorescent proteins or antibodies for bacterial detection purposes.

To reliably design tailor-made endolysins, knowledge about their characteristics, catalytic mechanisms, regulation is very advantageous. High-resolution structures of endolysins help to determine these properties. Determining the structure not only helps to elucidate the mechanism of action of that specific endolysin, but also allows to predict the structure of closely related enzymes.

Structure determination methods

Three main techniques can be used for structure determination: NMR spectroscopy, cryo-electron microscopy and X-ray crystallography (Mateu, 2013; Gauto et al*.,* 2019)**.** These techniques can be used alone or in combination (Everett et al*.,* 2015; Wang et al*.,* 2017).

Nuclear Magnetic Resonance (NMR) spectroscopy

Atomic nuclei are charged spinning particles. When exposed to a strong magnetic field and radiofrequency radiation, they will produce an electromagnetic signal with a frequency characteristic of the magnetic field at the nucleus. This signal will be different depending on the chemical environment, and the magnetic properties of every atom, making it useful for structure determination (Krishnan, 2019). The main advantages of NMR spectroscopy are that structures are determined in solution and the possibility of investigating protein dynamics and ligand binding properties. The main disadvantages are the requirement of high amounts of isotopically labelled sample and the fact that for large proteins and protein complexes, signals start to overlap, limiting the technique to smaller proteins (generally up to 30-50 kDa).

Cryo-Electron Microscopy (Cryo-EM)

In cryo-electron microscopy, small samples, in which the macromolecular particles are randomly oriented, are vitrified and kept at very low temperature. In the microscope, they are exposed to electrons to produce images of individual particles. These images are averaged to produce several two-dimensional classes that will be used to reconstruct the three-dimensional structure of the macromolecule. The main advantages of cryo-EM are the small amount of sample required and the fact that it is not necessary to obtain protein crystals. It is also suitable for large complexes, such as phage particles. On the other hand, cryo-EM does not work so well for smaller proteins (less than around 150 kDa). It is also not routinely possible to achieve atomic resolution structures.

X-ray crystallography

Crystallography uses diffraction to determine the position and arrangement of the atoms in a crystal. Using X-rays, atomic or near-atomic resolution can be achieved for small molecules, proteins and, in favourable cases, large complexes such as whole viruses. The main drawback of this technique is that crystals are necessary, and crystallisation of proteins requires high amounts of very pure and homogeneous sample and time. Often, a well-diffracting crystal can simply not be obtained. In addition to that, a static crystalline structure is obtained, although molecular dynamic techniques may provide clues about the

dynamics of the protein. Detailed protein-ligand interactions can be determined in favourable cases by co-crystallisation with a ligand or by soaking of ligands into the crystal. The latter is possible because protein crystals contain large solvent channels. Full-length endolysins vary in size. Their molecular weight may be too large to use NMR spectroscopy to determine their structure but too small to use cryo-EM. Due to this and to the nearly-atomic resolution obtained, X-ray crystallography is often the technique of choice.

Main steps in X-ray crystallography

The first step of X-ray crystallography is obtaining a protein sample with enough quality to be crystallised (Figure 1). This sample should be highly pure, stable, homogeneous, reproducible and abundant (Bergfors, 2009). If the protein contains several domains connected by flexible regions, the chances of getting protein crystals decrease. In these cases, truncated versions of the protein, i.e. removing several of the domains or limited proteolysis techniques to get rid of the flexible loops, are often useful.

After purification and concentration, the protein will remain in a soluble state as long as its concentration remains below its saturation point. In a supersaturated solution, protein molecules will tend to aggregate and, often, precipitate. To crystallise a protein, it is necessary to achieve protein supersaturation in a controlled way, causing the protein to aggregate in an ordered fashion, forming protein crystals. If supersaturation is reached slowly, precipitation will not occur immediately, and the protein will form nuclei. These nuclei are formed by a small number of protein molecules ordered in a lattice. Once these nuclei are formed, additional protein molecules will be incorporated in an ordered way, contributing to the crystal growth. The crystal will continue growing until the protein concentration decreases and gets below the supersaturation level.

Crystallisation is influenced by variables such as protein concentration, pH, temperature, and the use of different precipitants, salts and additives that can promote or delay protein aggregation (McPherson et al., 2014; Kirkwood et al., 2015). To crystallise a protein, different setups can be used. The most common is sitting-drop vapour diffusion. A drop containing a mix of the protein and the

crystallisation condition in different proportions sits on a pedestal next to a well or reservoir that contains a larger volume of the crystallisation condition. The system is covered by a non-permeable film. Due to the difference in concentration between the drop and the well, the solute of the drop will evaporate and will be transferred to the reservoir, causing the protein to slowly concentrate. If the conditions are optimal, supersaturation and controlled aggregation occur, leading to nucleation and crystal growth (Rupp, 2009).

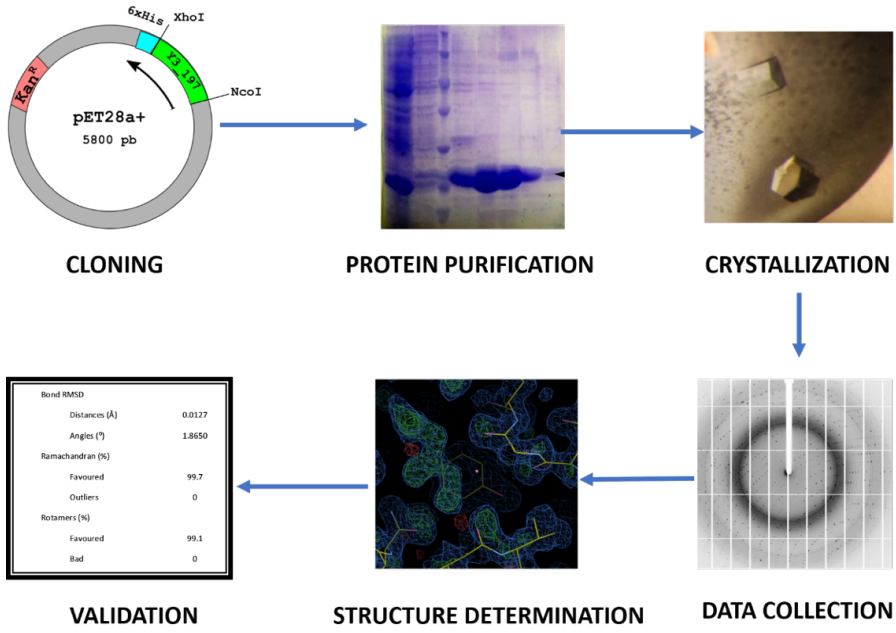


Figure 1: Main steps for protein structure determination using X-ray crystallography. See text for more details.

Once a good quality crystal is obtained, the three-dimensional structure of the protein forming it can be determined. This can be achieved by analysing the diffraction pattern generated by a highly parallel X-ray beam aimed at the crystal. After cryo-protecting and flash-cooling the crystal using liquid nitrogen, it

is exposed to a source of X-rays. X-rays are electromagnetic radiation with wavelengths ranging between 0.1 and 100 Å. Diffraction will occur if the wavelength of the radiation is similar to the spacing between the atoms of the lattice. In protein crystallography, the wavelengths most commonly used range from 0.6 Å to 2.5 Å, similar to the length of the chemical bonds in these molecules (Als-Nielsen et al*.,* 2011). Nowadays, the most commonly used X-ray sources are synchrotron radiation sources. Diffracted beams form a diffraction pattern, collected on a detector. The crystal is mounted on a goniometer head that rotates during X-ray data collection. That way, diffraction data can be collected at different angles, allowing for a complete diffraction data set.

X-rays are scattered from every point of the crystal, with a strength proportional to the concentration of electrons at that point. X-rays scattered along any particular direction interfere with each other and originate features in the diffraction pattern through constructive or destructive interference phenomena. For a large enough and well-ordered crystal, diffraction only occurs when Braggs' law is satisfied (Bragg, 1913). The spacing of the diffraction spots depends on the crystal symmetry and the size and shape of the crystal asymmetric unit (the smallest identifiable unique feature in the crystal that, when repeated, give rise to the complete, regular crystal). The intensity of the diffraction spots, however, depends on the internal structure of the asymmetric unit, which may contain one or more protein molecules.

When performing an X-ray crystallography experiment on a protein crystal, we obtain, for each spot in the diffraction pattern, the exact position and its intensity. Using this information, the symmetry of the crystal can be determined. Afterwards, the intensities of all the diffraction spots are determined (Leslie, 2006). The electron density in each asymmetric unit is related to the intensities of the diffraction spots by a mathematical formula known as the Fourier transform. Using this formula, if we know the exact content of the asymmetric unit (location and nature of each atom), we can predict the diffraction pattern. However, the inverse, and the goal of our experiment, i.e. determining the asymmetric unit contents from the diffraction pattern, is not that simple. To calculate the electron density within the asymmetric unit, we need not only the

intensities, but also the phase of the scattered X-rays (Taylor, 2010). Unfortunately, only the intensity of the diffraction spots can be measured experimentally, as the phase is lost. This is known as the phase problem (Cowtan, 2003).

To solve the phase problem, several approaches can be used (Taylor, 2010). In molecular replacement, we use the phases calculated from a protein model derived from a structure with high sequence similarity to ours (greater than 30% sequence identity as a rule of thumb), assuming their structures will be similar. For molecular replacement, first the model has to be placed in the correct location and orientation in the unit cell, using sophisticated software. In experimental phasing, a limited number of heavy atoms incorporated at specific places into the protein structure (for example mercury ions bound to accessible cysteine residues). This is usually done by trial and error soaking of heavy atom compounds into pre-formed crystals. A commonly used alternative method is to incorporate seleno-methionine residues into the protein by expressing the protein in special growth media and crystallising this seleno-methionine derivative version of the protein. The substructure of the heavy atoms is simpler and easier to calculate than the complete protein structure and can be used for phase determination.

Once the phases have been estimated, an electron density map of the protein can be obtained. A model that fits this data is then built, and if built correctly, this model then provides improved phases for calculation of an improved electron density map. The model is refined against the data, incorporating restraints limiting excessive variation of bond lengths, bond angles and other known protein structural parameters. As a final step, the structural model of the protein will be validated. The geometry of the final model will be analysed to detect potential outliers in rotamer prevalence, Ramachandran configurations, close contacts and other parameters that might have been introduced by the building or refinement process (Kleywegt, 2000). If the geometry of the model is correct, and the model to data statistics are reasonable, the model is deposited in the PDB, the Protein Data Bank (www.pdbe.org).

Endolysin structure

Catalytic domains of endolysins belong to one of the following enzyme families (Loessner, 2005; Fenton et al., 2010): N-acetylmuramidases (lysozymes), N-acetyl-β-D-glucosaminidases (glycosidases), N-acetylmuramoyl-L-alanine amidases, L-alanoyl-D-glutamate endopeptidases and interpeptide bridge-specific endopeptidases (Figure 2). Single-domain endolysins are exclusive to phages infecting Gram-negative bacteria and consist of just one catalytic domain. However, most endolysins exhibit a modular configuration with one or several catalytic domains and a cell wall binding domain. These domains are connected by flexible linkers. The lytic spectrum of an endolysin depends on its cell wall binding domain and on the conservation among bacteria of the cell wall motif that the cell wall binding domain recognises (Nelson et al., 2012, Navarre et al., 1999; O'Flaherty et al., 2005; Fernandes et al., 2018).

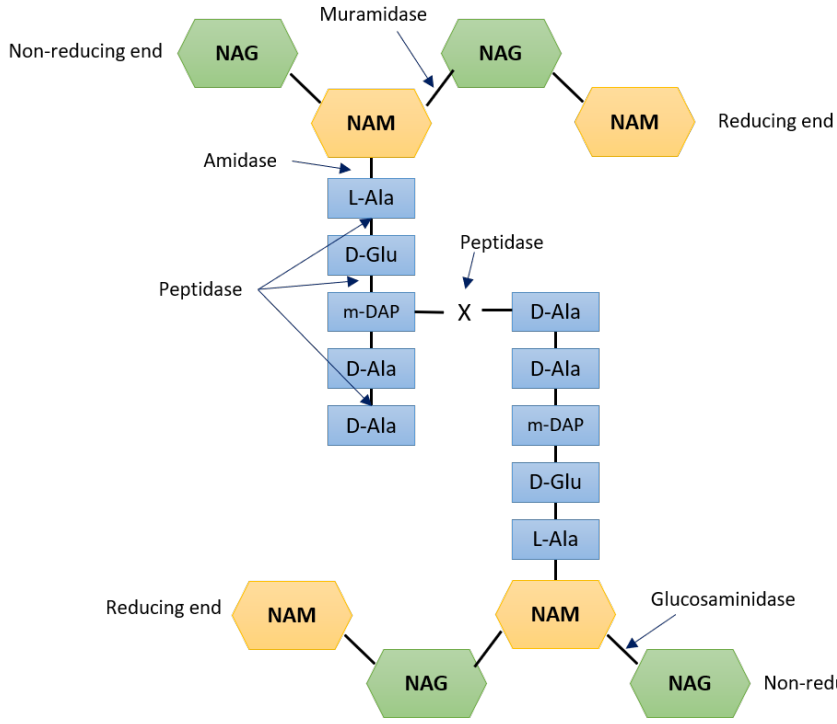


Figure 2: Enzymatic activities of peptidoglycan degrading enzymes. The basic structure of Gram-negative cell wall peptidoglycan, indicating the possible enzymatic activities of peptidoglycan-degrading enzymes and the bonds they cleave.

N-acetylglucosaminidases or glycosidases

N-acetylglucosaminidases hydrolyse the β-(1→4) glycosidic bond between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc). The only structure of a phage endolysin showing a N-acetylglycosidase domain is PlyC, the lysin from *Streptococcus* phage C1 (McGowan et al., 2012; Broendum et al., 2018). PlyC is a powerful multimeric endolysin that consists of two separate proteins: a single copy of the 50 kDa PlyCA subunit and at least eight copies of the 8 kDa PlyCB subunit. The eight PLyCB subunits form the cell wall binding domain of the protein and are arranged in a planar octameric ring 80 Å in diameter and 20 Å high. The PlyCA subunit contains two different domains, a carboxy-terminal cysteine, histidine-dependent amidohydrolase/peptidase (CHAP) domain and an amino-terminal N-acetylglycosidase domain. The N-acetylglycosidase domain is structurally related to N-acetylglucosaminidase domains from the glycosyl hydrolase 73 (GH73) family, with a highly conserved core formed by six α-helices. This core contains a conserved glutamic acid residue (glutamate 78) essential for catalytic activity and a substrate-binding groove (Figure 3A).

N-acetylmuramidases (lysozymes or muramidases)

N-acetylmuramidase domains hydrolyse the β-(1→4) glycosidic bond between N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) in the sugar moiety of the cell wall. Phage muramidase domains belong to at least three different families of glycosyl hydrolases called GH19, GH24 and GH25. Most belong to the GH25 family, including Cpl-1, the endolysin from the *Streptococcus pneumoniae* phage Cp-1 (Perez-Dorado et al., 2007). The Cpl-1 structure is formed by two domains connected to each other with a linker. The cell wall binding domain presents six choline-binding repeats. The muramidase domain forms an irregular $(\beta/\alpha)_5\beta_3$ barrel, in which the catalytic residues are located. All members of the GH25 family share this barrel fold. Hydrolysis in this group of muramidases takes place via a general acid/base mechanism that requires two acidic amino acids. One acidic residue behaves as the proton donor and the other acts as the nucleophile, promoting hydrolytic attack by a water molecule. These residues are located in a catalytic groove and are

strongly conserved. In the case of Cpl-1, the residues are glutamate 94 and aspartate 10 (Figure 3B, left).

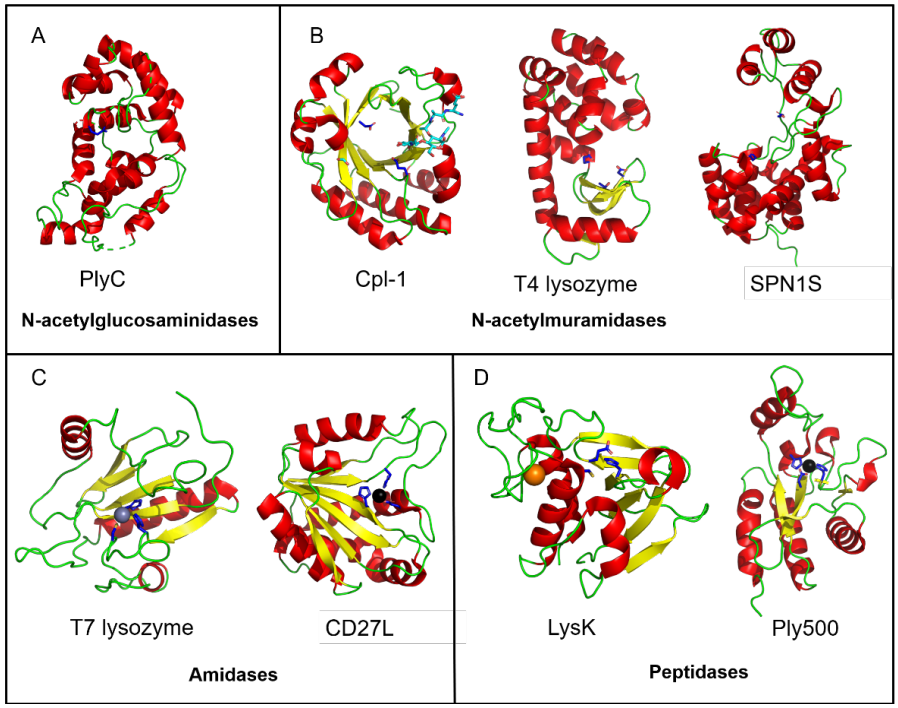


Figure 3. Representation of the variety of enzymatically active domains in phage endolysins. The proteins are shown in cartoon representation, coloured according to their secondary structure. α-helices are represented in red, β-sheets in yellow and loops in green. Carbon atoms of catalytic residues are depicted in dark blue, carbon atoms of ligands in cyan, zinc ions as dark grey or black spheres and calcium ions as orange spheres. **A.** N-acetylglucosaminidase domain from the streptococcal phage C1 lysin PlyC (PDB entry 4F88). **B.** Different families of N-acetylmuraminidases. Left: Cpl-1, endolysin from the *Streptococcus pneumoniae* phage Cp-1, representative of the GH25 family, (PDB entry 2IXU). Centre: T4 lysozyme, representative of the GH24 family (PDB entry 4LZM). Right: Endolysin from *Salmonella* phage SPN1S, representative of the GH19 family (PDB entry 4OK7). **C.** Different families of N-acetylmuramoyl-L-alanine amidases. Left: T7 lysozyme, representative of the Pfam amidase_2 family (PDB entry 1LBA). Right: CD27L, *Clostridium difficile* φCD27 phage endolysin, representative of the VanY d-Ala-d-Ala carboxypeptidase subfamily (PDB entry 3QAY). **D.** Different families of peptidases. Left: $CHAP_K$, CHAP domain of the staphylococcal phage K endolysin LysK (PDB entry 4CSH). Right: pLy500, L-alanoyl-D-glutamate peptidase from *Listeria* phage A500 (PDB entry 2VO9), representative of the M15C family of endopeptidases.

One of the most studied phage muramidases is T4 lysozyme (Bell et al*.,* 1991). T4 lysozyme has been studied in depth and exhaustive mutational analyses have been performed to infer the function of every residue (Rennell et al*.,* 1991). This enzyme is an inverting glycoside hydrolase that belongs to the GH24 family. T4 lysozyme is formed by two domains, that can move with respect to each other. The active-site cleft is located in the interface between the domains. Without the peptidoglycan substrate, T4 lysozyme remains in its open conformation. After peptidoglycan is added, the lysozyme transitions to a closed processive conformation. The change of conformation is caused by an 8 Å hinge motion that brings the two lobes of the enzyme closer to each other (Akhterov et al*.,* 2015). Once the enzyme is in a closed conformation, a nucleophilic water molecule, activated by aspartate 20, attacks the β-(1→4) linked anomeric C1 carbon of the peptidic N-acetyl muramic acid, while the general acid glutamate 11 donates a proton to the leaving aglycone, N-acetyl glucosamine (Kuroki et al*.,* 1995). Mediated by hydrogen bonds, threonine 26 helps to position the water molecule that performs the nucleophilic attack (Figure 3B, centre).

The endolysin from *Salmonella* phage SPN1S (Park et al*.,* 2014) belongs to the GH19 chitinase family, although it has muramidase activity. It is a helical monomer with a Pacman-like shape, formed by two domains: a large domain made up of the amino-terminal and carboxy-terminal regions of the protein and a smaller domain made up of the central part of the protein. The large domain is formed by a total of eight α-helices. Helix α3 is located in the central part of the domain, surrounded by helices α1, α2 and α10-α13. The small domain is formed by the five central α-helices (α4-α8) and a 3_{10}-helix. Helices α4, α6 and α7 form an antiparallel three-helix bundle whose axis is perpendicular to one of the bundles formed by the helices in the large domain. These two regions are connected by two long loops. As a result, a deep wide groove is formed between both domains, where the catalytic cleft is located. The catalytic residues from family 19 chitinases are two glutamates. A glutamate located at the end of the third α-helix (glutamate 49 in SPN1) acts as the catalytic general acid, whereas another glutamate located in a more variable loop-like structure (glutamate 58 in SPN1S) acts as the catalytic general base (Figure 3B, right).*N*-

Acetylmuramoyl-L-alanine amidases

Amidases cleave the amide bond connecting the sugar and peptide moieties of the bacterial cell wall. According to the classification performed by Broendum et al. (2018), two different zinc-dependent N-acetylmuramoyl-l-alanine amidase folds exist. The most common amidase domain belongs to the Pfam amidase_2 family (Firczuk et al., 2007). Although several phage endolysins contain this kind of fold, the more studied one has been T7 lysozyme (Cheng et al., 1994). These enzymes contain a five−stranded, mostly parallel β−sheet with conserved topology, flanked by a varying number of α−helices. The active site zinc ion is located on one of the sides of the β−sheet, at the bottom of the peptidoglycan binding groove. The zinc ion is tetrahedrally coordinated to a water molecule, two histidine residues and a cysteine residue. These residues are highly conserved (Figure 3C, left).

Other amidases, like the *Clostridium difficile* φCD27 phage endolysin CD27L (Mayer et al., 2011), belong to the VanY d-Ala-d-Ala carboxypeptidase subfamily. Despite belonging to this subfamily, they have N-acetylmuramoyl-L-alanine amidase activity. They exhibit an α/β fold, with a core of six beta strands surrounded by five alpha helices. A zinc ion is present in the catalytic groove, coordinated by two histidine residues and a glutamic acid. In the case of CD27L, these residues are histidine 9, histidine 84 and glutamate 26 (Figure 3C, right).

Peptidases

Peptidases cleave the peptide moiety of the peptidoglycan. Depending on the links in the peptidoglycan they target, there are two types, L-alanoyl-D-glutamate endopeptidases and interpeptide bridge-specific peptidases.

Most of the domains with peptidase activity found in phage endolysins are histidine-dependent amidohydrolase/peptidase (CHAP) domains. CHAP domains show a great versatility and can have N-acetylmuramoyl-L-alanine amidase or D-alanyl-glycyl endopeptidase activities (Bateman et al., 2003). CHAP domains are formed by two α-helices clustered against a six-stranded antiparallel β-sheet, accommodating a long and deep hydrophobic catalytic groove between them. The catalytic triad is formed by a cysteine, a histidine and

a glutamic acid residue, which are strongly conserved. Some CHAP domains, such as $CHAP_K$, the CHAP domain of the staphylococcal phage K endolysin LysK (Sanz-Gaitero et al*.,* 2014), contain a calcium ion bound to an EF-hand-like motif located in the amino-terminal part of the protein, involving residues from the long loop connecting the first and second α-helices (Figure 3D, left). This ion is necessary for the enzyme to be catalytically active. However, some other CHAP domains, like the one in pLyC, the lysin from the streptococcal phage C1, do not contain a calcium binding motif (McGowan et al*.,* 2012).

The only phage endolysin with endopeptidase activity that does not contain a CHAP domain is pLy500, the L-alanoyl-D-glutamate peptidase from *Listeria* phage A500 (Korndoerfer et al., 2008). Its catalytic domain belongs to the M15C family of endopeptidases. The overall shape of Ply500 catalytic domain resembles a sofa, with an α-helix and the three antiparallel β-strands forming the 'seat'. Another α-helix and a loop give rise to the 'backrest', and an additional loop conforms to the 'left armrest' of the sofa. A catalytically active zinc ion is tetrahedrally coordinated by a highly conserved set of residues (histidine 80, aspartate 87 and histidine 133) and a water molecule is located in a groove formed between the 'seat' and the 'backrest' (Figure 3D, right).

Cell wall binding domains

Cell wall binding domains are combined with enzymatically active domains to form modular endolysins. Cell wall binding domains bind non-covalently to a specific substrate (usually carbohydrate) found in the cell wall of the host bacterium, offering some degree of specificity to the endolysins. In general, while the catalytic domains of endolysins show a high sequence similarity among them, sequence similarity between cell wall binding domains is much lower (Fischetti, 2008). Cell wall binding domains show high affinity to their substrate, controlling diffusion of the enzymes and preventing the killing of neighbouring cells that have not yet produced phage particles (Loessner et al., 2002).

According to their structure, there are four main cell wall binding domain classes (Broendum et al., 2018):

1. SH3b domains, with a β-barrel fold, which can contain between five and seven β-strands connected by linkers. PlyPSA, the endolysin from the *Listeria* phage PSA (Korndoerfer et al., 2006) contains two SH3b repeats in its cell wall binding domain. These copies are held together by means of swapped β-strands (Figure 4A).
2. Choline binding modules (ChBr), often found in pneumococcal endolysins such as Cpl-1, which recognise the teichoic acids of the pneumococcal cell wall (Perez-Dorado et al., 2007). They are formed by multiple tandem copies of a short amino acid sequence. The Cpl-1 cell wall binding domain contains an amino-terminal super-helical subdomain formed by the first four ChBr repeats, where the choline-binding pockets are located, and a carboxy-terminal β-sheet that interacts with the catalytic domain (Figure 4B).
3. Three-helix bundles. In endolysins targeting Gram-positive bacteria, such as the *Streptococcus* phage CP-7 endolysin Cpl-7 (Bustamante et al*.,* 2017), several copies of this fold compose the cell wall binding domain. Endolysins targeting Gram-negative bacteria only contain one copy of the tree-helix bundle (Figure 4C).
4. α/β multimers. Several endolysin cell wall binding domains show a similar fold consisting of a central parallel or antiparallel four-stranded β-sheet flanked by two helices. One example of this kind of fold is PlyC, the lysin from the streptococcal phage C1 (McGowan et al., 2012). These cell wall binding domains oligomerize to form a functional assembly (Figure 4D).

Endolysin variability

Phage-encoded endolysins show a lot of diversity. They can contain a single globular catalytically active domain or combine different cell wall binding domains and enzymatically active domains to form a multi-modular endolysin. This flexibility in their design makes phage endolysins highly versatile proteins, with different catalytic activities, host ranges and stabilities. A comparative genomic analysis of 723 endolysins identified 24 different enzymatically active domains and 13 cell wall binding domain types arranged in 89 different architectures, highlighting the high variety among phage endolysins (Oliveira et al*.,* 2013).

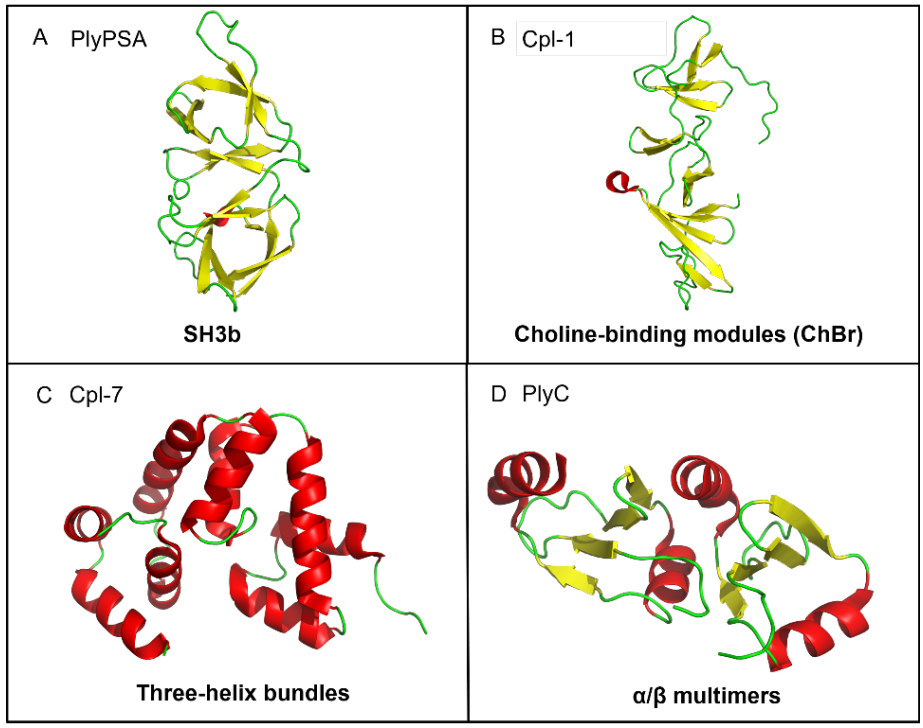


Figure 4. Representation of the variability of cell wall binding domains in phage endolysins. The proteins are shown in cartoon representation, coloured according to their secondary structure. α-helices are represented in red, β-sheets in yellow and loops in green. **A.** PlyPSA, endolysin from the *Listeria* phage PSA (PDB entry 1XOV) contains two SH3b copies in its cell wall binding domain. **B.** Choline binding module (ChBr) in the cell wall binding domain of Cpl-1, the endolysin from the *Streptococcus pneumoniae* phage Cp-1 (PDB entry 2IXU). **C.** *Streptococcus* phage CP-7 endolysin Cpl-7 cell wall binding domain showing three copies of a three-helix bundle (PDB entry 5I8L). **D.** Dimer of PlyCB, the cell wall binding domain of the endolysin from the *Streptococcus* phage C1 (PDB entry 4F87), forming α/β multimers.

The multi-modular nature of many endolysins offers the opportunity to design tailor-made proteins with properties adapted for different applications, by combining, truncating or mutating these domains. Due to the flexibility between modules, not many full-length multi-modular endolysins structures have been published. A common approach in structure determination of multi-modular

endolysins is to study truncated versions of the endolysin, containing just one of the stable domains.

Some endolysins are formed by more than one peptide chain. Staphylococcal phage C1 endolysin, PlyC, is formed by single copy of the 50 kDa PlyCA subunit and at least eight copies of the 8 kDa PlyCB subunit (see the section *N-acetylglucosaminidases*). *Enterococcus* phage IMEEF1 endolysin LysIME-EF1 (Zhou et al., 1994) is also a two-component endolysin. The active protein is a hetero-tetramer formed by one full-length LysIME-EF1 copy and three additional carboxy-terminal cell wall binding domains. The cell wall binding domains are truncated versions of the full-length protein translated from an alternative ribosomal binding site within the gene.

Clostridium difficile phage φCD27 endolysin CD27L (Dunne et al., 2014) contains an amino-terminal amidase domain and a carboxy-terminal cell wall binding domain. A self-cleavage site is located in the linker between both domains. CD27L exists in two different dimeric states. The tensed inactive state is a head-on dimer, where the two autocleavage sites are far away from each other. The relaxed active state is the side-by-side dimer, which promotes autocleavage and the release of the amino-terminal catalytic domain from the carboxy-terminal cell wall binding domain. Autocleavage increases the action radius of the catalytic module and may allow it to penetrate the bacterial cell wall independently of the cell wall binding domain.

The structures from two different 'SAR' endolysins have been solved using X-ray crystallography, Lyz of coliphage P1 (Xu et al., 2005) and R21 of the lambdoid phage 21 (Sun et al., 2009). These enzymes contain an amino-terminal SAR (Signal-Anchor-Release) domain that allows the translocation to the periplasm through the Sec system, in a holin-independent manner. This SAR domain is not cleaved by the signal peptidase after the translocation, and the SAR endolysin remains attached to the membrane in its inactive conformation during phage maturation thanks to the amino-terminal transmembrane domain embedded in the inner membrane. To become active, the endolysin must be released from the inner membrane and refolded into the catalytically active

form. Dissipation of the proton motive force causes the transmembrane domain to be released from the cell membrane. After leaving the bilayer, the protein refolds into a periplasmic active form (Fernandes et al*.,* 2018). The holins facilitate this process by mediating the collapse of the proton motive force (Xu et al*.,* 2004; Xu et al*.,* 2005; Sun et al*.,* 2009).

Conclusion

X-ray crystallography has contributed significantly towards the knowledge of endolysin structure and mechanism. Structures of catalytic domains have provided clues as to their enzymatic mechanisms, while structures of cell wall binding domains have informed us as to how they specifically recognise the peptidoglycan of their bacterial hosts. However, structures of endolysin domains complexed with peptidoglycan substrates are still scarce. Future improvements in the organic synthesis of peptidoglycan analogues will hopefully fill this void. In any case, structural biology in general, and X-ray crystallography in particular, will continue to play a role in determining structures of endolysins, endolysin domains and their complexes with substrate analogues. In turn, this structural information will guide the design of more stable and more efficient enzybiotics to eradicate pathogenic bacteria and also positively influence the use of endolysin-based bacterial detection methods.

Acknowledgements

This work was supported by grant BFU2017-82207-P from the Spanish Ministry of Science and Innovation to MJvR and by a RISAM PhD fellowship from the Cork Institute of Technology to MS-G.

References

Akhterov, M.V., Choi, Y., Olsen, T.J., Sims, P.C., Iftikhar, M., Gul, O.T., Corso, B.L., Weiss, G.A. and Collins, P.G. (2015) Observing lysozyme's closing and opening motions by high-resolution single-molecule enzymology. ACS Chem. Biol. 10, 1495–1501. https://doi.org/10.1021/cb500750v.

Alcorlo, M., Martinez-Caballero, S., Molina, R. and Hermoso, J.A. (2017) Carbohydrate recognition and lysis by bacterial peptidoglycan hydrolases. Curr. Opin. Struct. Biol. 44, 87–100. https://doi.org/10.1016/j.sbi.2017.01.001.

Als-Nielsen, J. and McMorrow, D. (2011) Elements of modern X-ray physics. (New York, USA: John Wiley & Sons). https://doi.org/10.1002/9781119998365.

Bateman A., Rawlings N.D. (2003) The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends. Biochem. Sci. 28, 234–7. https://doi.org/10.1016/S0968-0004(03)00061-6.

Bell, J. A., Wilson, K. P., Zhang, X.J., Faber, H. R., Nicholson, H. and Matthews, B. W. (1991) Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. Proteins 10, 10–21. https://doi.org/10.1002/prot.340100103.

Bergfors, T. (2009) Protein crystallization. (La Jolla, California, USA: International University Line).

Bragg, W.L. (1913) The diffraction of short electromagnetic waves by a crystal, Proceedings of the Cambridge Philosophical Society, 17, 43–57.

Broendum, S.S., Buckle, A. M. and McGowan, S. (2018) Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins. Mol. Microbiol. 110, 879–896. https://doi.org/10.1111/mmi.14134.

Bustamante, N., Iglesias-Bexiga, M., Bernardo-García, N., Silva-Martín, N., García, G., Campanero-Rhodes, M.A., et al. (2017) Deciphering how Cpl-7 cell wall-binding repeats recognize the bacterial peptidoglycan. Sci. Rep., 7, 16494. https://doi.org/10.1038/s41598-017-16392-4.

Cheng X., Zhang X., Pflugrath J.W., Studier F.W. (1994) The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. 91 ,4034-4038. https://doi.org/10.1073/pnas.91.9.4034.

Coffey, A. and Buttimer, C. (2020). Bacterial Viruses: Exploitation for Biocontrol and Therapeutics (Norfolk, UK: Caister Academic Press). https://doi.org/10.21775/9781913652517

Cowtan K (2003) Phase problem in X-ray crystallography, and its solution. In Encyclopedia of life sciences. (Chichester, UK: Wiley). https://doi.org/10.1038/npg.els.0002722.

Dunne, M., Mertens, H. D., Garefalaki, V., Jeffries, C. M., Thompson, A., Lemke, E. A., Svergun, D. I., Mayer, M. J., Narbad, A., and Meijers, R. (2014). The CD27L and CTP1L endolysins targeting Clostridia contain a built-in trigger and

release factor. PLoS Pathog. 10, e1004228. https://doi.org/10.1371/journal.ppat.1004228

Everett, J.K., Tejero, R., Murthy, S.B., Acton, T.B., Aramini, J.M., Baran, M.C., Benach, J., Cort, J.R., Eletsky, A., Forouhar, F., Guan, R., Kuzin, A.P., Lee, H.W., Liu, G., Mani, R., Mao, B., Mills, J.L., Montelione, A.F., Pederson, K., Powers, R., et. al. (2016) A community resource of experimental data for NMR / X-ray crystal structure pairs. Protein Sci. 25, 30–45. https://doi.org/10.1002/pro.2774.

Fenton, M., Ross, P., Mcauliffe, O., O'Mahony, J. and Coffey, A. (2010) Recombinant bacteriophage lysins as antibacterials. Bioeng. Bugs. 1, 9-16. https://doi.org/10.4161/bbug.1.1.9818.

Firczuk, M. and Bochtler, M. (2007) Folds and activities of peptidoglycan amidases. FEMS Microbiol. Revi. 31, 676–69. https://doi.org/10.1111/j.1574-6976.2007.00084.x.

Fischetti, V. A. (2008) Bacteriophage lysins as effective antibacterials. Curr. Opin. Microbiol. 11, 393-400. https://doi.org/10.1016/j.tim.2005.08.007.

Gauto, D.F., Estrozi, L.F., Schwieters, C.D., Effantin, G., Macek, P., Sounier, R., Sivertsen, A.C., Schmidt, E., Kerfah, R., Mas, G., Colletier, J.P., Güntert, P., Favier, A., Schoehn, G., Schanda, P and Boisbouvier, J. (2019). Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. Nat. Comm. 10, 2697. https://doi.org/10.1038/s41467-019-10490-9.

Gerstmans, H., Criel, B. and Briers, Y. (2018) Synthetic biology of modular endolysins. Biotechnol. Adv. 36, 624–640. https://doi.org/10.1016/j.biotechadv.2017.12.009.

Kakasis, A. and Panitsa G. (2019) Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. Int. J. Antimicrob. Agents. 53, 16-21. https://doi.org/10.1016/j.ijantimicag.2018.09.004.

Kirkwood, J., Hargreaves, D., O'Keefe, S. and Wilson, J. (2015) Analysis of crystallization data in the Protein Data Bank. Acta. Cryst. F71, 1228–1234. https://doi.org/10.1107/S2053230X15014892.

Kleywegt, G.J. (2000) Validation of protein crystal structures. Acta. Cryst. D56, 249–265. https://doi.org/10.1107/S0907444999016364.

Korndoerfer, I.P., Danzer, J., Schmelcher, M., Zimmer, M., Skerra, A. and Loessner, M.J. (2006) The crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for specific recognition of *Listeria* cell walls. J. Mol. Biol. 364, 678–689. https://doi.org/10.1016/j.jmb.2006.08.069.

Korndoerfer, I.P, Kanitz, A., Danzer, J., Zimmer, M, Loessner, M.J., Skerra, A. (2008) Structural analysis of the l-alanoyl-d-glutamate endopeptidase domain of Listeria bacteriophage endolysin Ply500 reveals a new member of the LAS peptidase family. Acta Cryst. D64, 644–650. https://doi.org/10.1107/S0907444908007890.

Krishnan V.V. (2019) Molecular thermodynamics using nuclear magnetic resonance (NMR) spectroscopy. Inventions *4*, 13. https://doi.org/10.3390/inventions4010013.

Kuroki, R., Weaver, L.H. and Matthews, B.W. (1995) Structure-based design of a lysozyme with altered catalytic activity. Nat. Struct. Biol. 2 ,1007-1011. https://doi.org/10.1038/nsb1195-1007.

Leslie, A.G. (2006) The integration of macromolecular diffraction data. Acta Cryst. D62, 48–57. https://doi.org/10.1107/S0907444905039107.

Loessner, M.J. (2005) Bacteriophage endolysins-current state of research and applications. Curr. Opin. Microbiol. 8, 480-487. https://doi.org/10.1016/j.mib.2005.06.002.

Loessner, M.J., Kramer, K., Ebel, F. and Scherer, S. (2002) C-terminal domains of Listeria monocytogenes bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. Mol. Microbiol. 335-349. https://doi.org/10.1046/j.1365-2958.2002.02889.x.

Mateu M.G. (2013) Introduction: The Structural Basis of Virus Function. In Structure and Physics of Viruses. Subcellular Biochemistry. M.G Mateu. ed. (Dordrecht, Netherlands: Springer) vol 68, pp. 3-51. https://doi.org/10.1007/978-94-007-6552-8_1.

Majewska, J., Kaźmierczak, Z., Lahutta, K., Lecion, D., Szymczak, A., Miernikiewicz, P., Drapała, J., Harhala, M., Marek-Bukowiec, K., Jędruchniewicz, N., Owczarek, B., Górski, A., and Dąbrowska, K. (2019). Induction of phage-specific antibodies by two therapeutic staphylococcal

bacteriophages administered *per os*. Front. Immunol., 14 November 2019. doi: 10.3389/fimmu.2019.02607

Mayer, M.J., Garefalaki, V., Spoerl, R., Narbad, A., and Meijers, R. (2011) Structure-based modification of a Clostridium difficile-targeting endolysin affects activity and host range. J. Bacteriol. 193, 5477–5486. https://doi.org/10.1128/JB.00439-11.

McGowan, S., Buckle, A.M., Mitchell, M.S., Hoopes, J.T., Gallagher, D.T., Heselpoth, R.D., et al. (2012) X-ray crystal structure of the streptococcal specific phage lysin PlyC. Proc. Natl. Acad. Sci. U. S. A. 109, 12752–12757. https://doi.org/10.1073/pnas.1208424109.

McPherson, A. and Cudney, B. (2014) Optimization of crystallization conditions for biological macromolecules. Acta. Cryst. F70, 1445–1467. https://doi.org/10.1107/S2053230X14019670.

Navarre, W.W., Ton-That, H., Faull, K.F. and Schneewind, O. (1999) Multiple enzymatic activities of the murein hydrolase from staphylococcal phage φ11. Identification of a D-alanyl-glycine endopeptidase activity. J. Biol. Chem., 274, 15847-15856. https://doi.org/10.1074/jbc.274.22.15847.

Nelson, D.C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D.G., Dong, S. and Donovan, D.M. (2012) Endolysins as antimicrobials. Adv. Virus. Res. 83, 299–365. https://doi.org/10.1016/B978-0-12-394438-2.00007-4.

O'Flaherty, S., Coffey, A., Meaney, W., Fitzgerald, G.F. and Ross, R.P. (2005) The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant *Staphylococcus aureus*. J. Bacteriol. 187, 7161-1764. http://doi.org/ 10.1128/JB.187.20.7161-7164.2005.

O'Sullivan, L., Buttimer, C., McAuliffe, O., Bolton, D. and Coffey, A. (2016) Bacteriophage-based tools: recent advances and novel applications. F1000Research 5, 2782. https://doi.org/10.12688/f1000research.9705.1.

Oliveira H., Melo L.D., Santos S.B., Nóbrega F.L., Ferreira E.C., Cerca N., Azeredo J. and Kluskens L.D. (2013) Molecular aspects and comparative genomics of bacteriophage endolysins. J. Virol. 87, 4558–4570. https://doi.org/10.1128/JVI.03277-12.

Park, Y., Lim, J.A, Kong, M., Ryu, S. and Rhee, S. (2014) Structure of bacteriophage SPN1S endolysin reveals an unusual two-module fold for the

peptidoglycan lytic and binding activity. Mol. Microbiol. 92, 316–325. https://doi.org/10.1111/mmi.12555.

Pérez-Dorado, I., Campillo, N.E., Monterroso, B., Hesek, D., Lee, M., Páez, J.A., et al. (2007) Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. J. Biol. Chem., 282, 24990–24999. https://doi.org/10.1074/jbc.M704317200.

Rennell, D., Bouvier, S.E., Hardy, L.W. and Poteete, A.R. (1991) Systematic mutation of bacteriophage T4 lysozyme. J. Mo.l Biol. 222, 67–88. https://doi.org/10.1016/0022-2836(91)90738-R.

Rupp, B. (2009) Biomolecular crystallography: principles, practice and application to structural biology (New York, USA:Garland Science).

Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A., and van Raaij, M.J. (2014) Crystal structure of the lytic CHAPk domain of the endolysin LysK from *Staphylococcus aureus* bacteriophage K. Virol. J. 11, 133. https://doi.org/10.1186/1743-422X-11-133.

Sun, Q., Kuty, G.F., Arockiasamy, A., Xu, M., Young, R. and Sacchettini, J.C. (2009) Regulation of a muralytic enzyme by dynamic membrane topology. Nat Struct. Mol. Biol. 16, 1192–1194. https://doi.org/10.1038/nsmb.1681.

Taylor, G.L. (2010) Introduction to phasing. Acta Cryst. D66, 325–328. https://doi.org/10.1107/S0907444910006694.

Wang, H.W. and Wang, J.W. (2017). How cryo-electron microscopy and X-ray crystallography complement each other. Protein Sci. 26, 32–39. https://doi.org/10.1002/pro.3022.

Xu, M., Arulandu, A., Struck, D. K., Swanson, S., Sacchettini, J. C. and Young, R. (2005) Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science 307, 113–117. https://doi.org/10.1126/science.1105143.

Xu, M., Struck, D.K., Deaton, J., Wang, I.N. and Young, R. (2004) The signal arrest-release (SAR) sequence mediates export and control of the phage P1 endolysin. Proc. Natl. Acad. Sci. U. S. A. 101, 6415–6420. https://doi.org/10.1073/pnas.0400957101.

Yang, H., Yu, J. and Wei, H. (2014) Engineered bacteriophage lysins as novel anti-infectives. Front. Microbiol. 5, 1–6. https://doi.org/10.3389/fmicb.2014.00542.

Young, R. (2013) Phage lysis: Do we have the hole story yet? Curr. Opin. Microbiol. 16, 790–797. https://doi.org/10.1016/j.mib.2013.08.008.

Zhou, B., Zhen, X., Zhou, H., Zhao, F., Fan, C., Perčulija, V., Tong, Y., Mi, Z., and Ouyang, S. (2020). Structural and functional insights into a novel two-component endolysin encoded by a single gene in *Enterococcus faecalis* phage. PLoS pathogens 16, e1008394. https://doi.org/10.1371/journal.ppat.1008394.